

中間概念としての認識構造の記述形式

川辺 諭

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

時枝誠記の言語理論を発展的に継承した三浦つとむの文法理論 [1][2][3] では、言語表現を客観的な対象世界のあり方を述べる客体表現と、話者の加える主体的な判断である主体表現に分離している。また 言語表現は、“対象－認識－表現”といった過程的な構造において行われるとしている(言語過程説)。

本稿では多言語翻訳において、言語過程説における認識構造を計算機内に記述し、中間概念として運用することを試みる。英語基本語彙として中学校で学習する英単語を対象に、認識構造の記述方法を検討し、中間概念として認識構造を用いる利点を述べる。

2 認識構造の記述形式

認識構造は動詞と格要素を列挙した格リストと、時制や様相などを列挙した機能リストからなる。認識構造の例を以下に挙げる。

[認識構造,
[格リスト,
[動的概念,\$授与],
[動作主格,\$人_1 人称],
[対格,\$具体物_本],
[与格,\$人_3 人称,\$性別, 男]],
[場所格,\$施設_学校]],
[機能リスト,
[様相, 肯定],
[時制, 現在]]]

英語表現: I give a book to him at school.

図 1: 認識構造の例

\$で記述される要素は中間概念である。以下に格リストと機能リストを構成する要素を述べる。

2.1 格リスト

格リストには動的属性と格要素が列挙される。

2.1.1 動的属性

動的属性は対象世界の事象を表現する。内部データとして、事象に関与する実体を列挙した格フレームを準備する。格フレームには実体の格やクラス、インスタンス情報と、実体の属性変化が記述される。格フレームの例を以下に示す。この例では対格が、\$取得といった動的概念によって動作主格の内部に取り入れられるといった内容が記述されている。

[動的概念, \$取得,

[動作主格, \$動物],

[対格, \$食物,

[変化属性, [\$場所, ~内部 (動作主格), 内部 (動作主格)]]]

対応語彙の例: 英語動詞 eat

図 2: 動的概念格フレームの例

翻訳対象となる言語の語彙とこれらの中間概念が、関係づけられる。

またこれらの動的属性は、格要素とその属性に着目し、類義のものを集め、継承関係を設けることで多次元シンソーラスを構築する(図 3)。

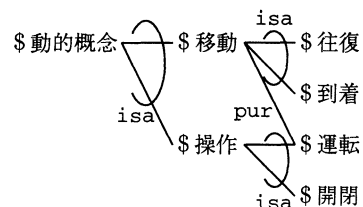


図 3: 動的概念多次元シンソーラスの例

多次元ソーラスでは、1つのノードを複数ノードの下位概念として扱うことができる。図3では\$運転といった動的属性を、“動作主体が移動する”といった観点から\$移動の下位概念として、また“人工物を操作する”といった観点から\$操作の下位概念として関係づけられている。

このように動的属性は、表現上を用いる格要素とその属性変化に着目することで、多様な観点から分類することが可能である。多次元ソーラスでは語彙や中間概念のデータを、言語処理上有用な観点から分類する。

2.1.2 実体

実体は言語表現では名詞で表され、属性の集合として記述される。実体の特定の属性に着目することで、動的概念と同様に多次元ソーラスを構築する(図4)。

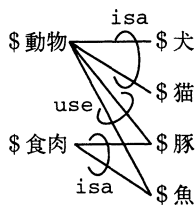


図4：実体多次元ソーラスの例

\$豚や\$魚は動物、食肉といった多面性を持っており、言語処理を行う文脈、状況に応じて異なった属性が着目される。多次元ソーラスでは\$動物、\$食肉といった複数の上位概念と関係づけることで、これらの問題に柔軟に対応する。

多次元ソーラスを構築する時の分類観点の例を以下に挙げる。

- isa
isa 関係では、上位概念にさらに属性を付加する(意味的編入)ことにより、詳細化された下位概念が得られる。上位概念として\$乗り物、下位概念として\$バス、\$自転車といった関係が挙げられる。
- hasa
hasa 関係は、下位概念が上位概念の構成要素となる関係である。上位概念として\$人、下位概念として\$耳、\$目といった関係が挙げられる。
- func
\$人工物において言語運用上最も着目される属性は

\$機能である。\$車や\$自転車は\$移動という機能といった観点から束ねると、言語処理上有用である。

- use
\$自然物の人間にとっての利用法である。\$食肉と\$魚といった関係が挙げられる。
- pur
\$動的属性と、それを達成するための\$動的属性の関係である。\$移動と\$運転といった関係が挙げられる。

2.1.3 静的属性

属性は、認識構造を構成する最もプリミティブな要素であり、細分化されることはない。ソーラスを構築して管理しているが、名詞や動詞の多次元ソーラスのような多重継承は行わない。属性の記述方法は以下の2通りである。

- 属性+属性値
属性値が絶対的、相対的な値で表現されるものである。[大小, 大]といった物理的属性や、[快樂, 大]といった心理的属性などがある。
- 属性+テンプレート
属性値にテンプレートを用いるものである。[形状, 球]や[色合い, 赤]といったものがある。

2.2 機能リスト

機能リストに記述される内容は以下のものである。

- 機能：叙述、疑問など
- 時制：現在、過去、未来など。
- 相：進行、完了、継続、開始など。
- 様相：肯定、否定、蓋然、許可、義務、命令、依頼、勧誘など。

3 認識構造を用いた機械翻訳

3.1 言語表現と認識構造

原言語を構文解析し、語彙に対応する中間概念を適用することで認識構造が得られる。またこれらの認識

構造の中間概念に目標言語の語彙を適用することで、目標言語表現を生成する。

** 言語表現 **

英語：I watched the movie.

日本語：私は映画を見た。

中国語：我看电影了。

** 認識構造 **

[認識構造,

[格リスト,

[動的概念,\$取得_情報],

[動作主格,\$人_1 人称],

[対格,\$抽象物_情報_映画]],

[機能リスト,

[様相, 肯定],

[時制, 過去]]]\$

図5：言語表現と認識構造

3.2 リボルバー変換

実際の機械翻訳では、言語間において事象の捉え方が違うため、認識構造のあり方が大きく異なることがある。

例えば日本語と英語では、知人や家族など特定の関係を持った人間が、身のまわりにいることを表現する方法が異なる。

日本語：私には娘がいる。

英語：I have a daughter.

この例では一方は「娘」を\$存在といった動的概念で捉え、他方は“a daughter”を動作主格の\$所有といった動的概念で捉えている。そのため、認識構造のあり方も大きく違う。図6に、認識構造の格リストの部分だけを抜き出す。

** 日本語側 **

[格リスト,

[動的概念,\$存在],

[動作主格,\$人_家族_娘],

[場所格,\$所有空間,\$人_1 人称]]]

** 英語側 **

[格リスト,

[動的概念,\$所有],

[動作主格,\$人_1 人称],

[対格,\$人_家族_娘]]]\$

図6：2つの認識構造中の格リストの違い

各々の言語におけるこれらの認識構造の違いを吸収するために、同一の事象をさまざまな観点から捉えた時の認識構造の差分を、変換対としてまとめておき、目標言語の性質によってそれらを適宜選択する。この手法をリボルバー変換と呼んでいる。

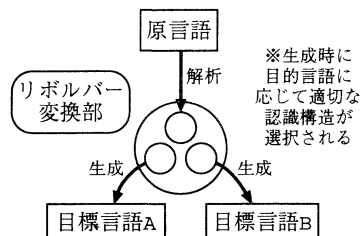


図7：リボルバー変換の概念図

3.3 リボルバー変換対の取得

言語間におけるリボルバー変換対を取得するために、既存の電子された言語コーパスや、和英、英和辞典などの対訳対を利用する。

対訳文のそれぞれに対して形態素、構文解析を行ない認識構造を作成し、その共通部分と差分を半自動抽出する。抽出された差分の集合がリボルバー変換対となる(図8)。

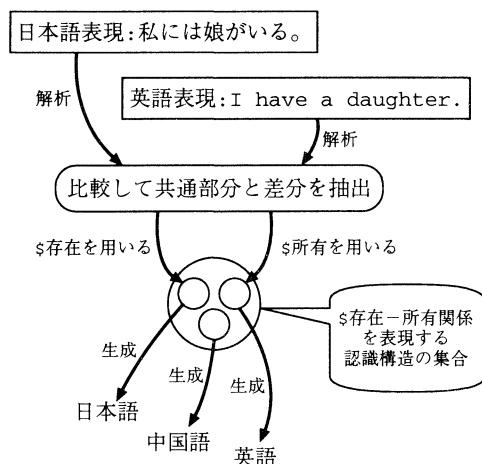


図8：リボルバー変換対の取得

3.4 リボルバー変換の例

3.4.1 イベントの表現

\$催し物が控えているという容の表現は、日本語と英語の間で以下のように異なる。

日本語： 今日パーティがある。

英語： We have a party today.

一方は\$催し物を\$存在という動的概念で、他方は\$催し物を\$所有という動的概念で表現している。

3.4.2 無生物主語

無生物主語など、構文の違いもリボルバー変換部で吸収する。以下に例を挙げる。

日本語： 英語を話すのは難しい。

英語： It is hard to speak English.

To speak English is hard.

Speaking English is hard.

3.4.3 日本語サ変動詞における対格の変換

日本語サ変動詞では、

「<対格>を<サ変動詞>する」

(部屋を掃除する)

「<対格>の<サ変動詞>をする」

(部屋の掃除をする)

の双方の表現が使われる。このような各言語の局所的な変換も、リボルバー変換で対応する。

4 おわりに

多言語翻訳における中間概念として時枝・三浦の言語過程説における認識構造を用いる方法を提案した。中間概念として認識構造を用いた多言語翻訳では、特定の言語対毎の解析部、生成部を作成する必要がなく、実装の手間が省ける。また、リボルバー変換部を準備することで、各言語における対象世界の捉えかたの違いによる、認識構造の違いに柔軟に対応できる。

今後の課題としては

- 本機構の実装をすすめる

- 言語データ拡充

- 本機構の有用性の検討

などがあげられる。

参考文献

- [1] 三浦つとむ:認識と言語の理論, 第1部～第3部, 勁草書房(1967/1967/1972).
- [2] 三浦つとむ:日本語の文法, 勁草書房(1975).
- [3] 三浦つとむ:日本語とはどういう言語か, 講談社学術文庫(1976).
- [4] 田中茂範:認知意味論 英語基本動詞の多義構造, 三友社出版(1990).
- [5] 田中茂範:基本動詞の意味論, 三友社出版(1987).
- [6] 川村, 片桐, 宮崎:語を種々の観点から分類した多次元シソーラス, 信学技法, NLC94-48, pp.33-40(1995).
- [7] 中田, 宮崎:類語弁別ネットワークによる語義記述法, 自然言語処理の基本問題シンポジウム論文集, pp.1-8(1992).